RELATIONSHIP BETWEEN CLIMATE AND THE MEAN ANNUAL FLOW OF THE MISSISSIPPI RIVER AT ST. PAUL

SPECIAL REPORT NO. 11

RICHARD H. SKAGGS
DWIGHT A. BROWN

DEPARTMENT OF GEOGRAPHY

WATER RESOURCES RESEARCH CENTER
UNIVERSITY OF MINNESOTA
ST. PAUL, MN 55108
JUNE 1987

RELATIONSHIP BETWEEN CLIMATE AND THE MEAN ANNUAL FLOW OF THE MISSISSIPPI RIVER AT ST. PAUL

Richard H. Skaggs
Dwight A. Brown

ABSTRACT

This study demonstrates the use of simulation of water balance for large areas. It examines the statistical relationship between the mean annual flow of the Mississippi River at St. Paul and the water balance surpluses (water not used in evapotranspiration and soil moisture storage) for the six climatological divisions of Minnesota that contribute to the drainage of the Mississippi and Minnesota Rivers in Minnesota. The water surpluses are calculated by the Thornthwaite method of estimating the water balance.

About 70 percent of the mean annual flow of the Mississippi River at St. Paul is statistically explained by variations in the surpluses computed for these two divisions.

Although the statistical model was developed on the first half of an eighty three year record of mean annual flows (1900-1982) it was quite capable of predicting mean annual flows of the second half of the record. From these results we conclude that it possible to statistical predict large scale water resources in Minnesota if the data are collected very rapidly (near real time) and used in a water budget calculation (such as the Thornthwaite method) to determine the water surpluses.

From the excellent performance of the statistical equations in the second half of the record we conclude that a substantial component of the increase in mean annual flow of the Mississippi in the past 40 years results from greater precipitation since 1940 in both the eastern and western portions of Minnesota and not lower temperatures.

We followed the development of a statistical model with a simulation of the possible effects of a temperature increase, resulting from the "Greenhouse Effect" (increasing carbon dioxide in the atmosphere). We used a postulated a 3 degree Celsius temperature rise and recomputed the water surpluses for the west central and east central climatological divisions. The results suggest a major decrease in the amount of surplus in both divisions if the temperature rise occurs and a corresponding reduction in the mean flow of the Mississippi River at St. Paul.

The approach used here assumes uniform soil and land-cover. It is presented as an example of the kinds of atmospheric and surface water resource questions that are appropriately addressed by simulation methods. Good soils and land-cover data stored in GIS files should provide greater flexibility in formulating simulation problems and even more defendable simulation results for areas that have the type of used here.

RELATIONSHIP BETWEEN CLIMATE AND THE MEAN ANNUAL FLOW OF THE MISSISSIPPI RIVER AT ST. PAUL

Richard H. Skaggs
Dwight A. Brown

INTRODUCTION

The magnitude of Minnesota's water resources varies at all time and space scales. Individual wells on farms and suburban lots, small watersheds, and the major river systems constantly alternate between deficit and surplus. Seldom is the current value of the resource equal to the long term mean. To design an adequate method of modeling these fluctuations in near real time and forecasting, stochastically, future fluctuations, it necessary to establish some of the links between the size of the water resource and the factors that control the magnitude.

In this report we explore the apparent links between the contemporary water resource and climate for a large area. We do this by using a fitted model of atmospheric moisture surplus to develop a statistical model of the relationships between atmospheric moisture surplus and river flow. This is done as a demonstration of how the fitted water budget models can be linked with statistical models to simulate the effect of carbon dioxide increases on surface water resources. Such simulation exercises can help us to understand future outcomes and possibly cope with the causes of problems rather than react to to the consequences. We use a scenario of the future in which CO2 is doubled. We selected the mean annual flow of the Mississippi/Minnesota drainage at St. Paul as a crude but robust and easily available measure of the available water in a large system. Because the area drained by the Mississippi at St. Paul is a large fraction of the State, it is unlikely that climatic data for one or even many individual climate observing stations are adequate to describe the relationship between the size of the flow in the Mississippi and the climate. Thus, we choose to use climatological divisions as the base climate data.

Of the 9 climatological divisions of Minnesota, 6 contribute to the drainage of the Mississippi at St. Paul (Figure 1). Our estimates of percentage of the division area that drains to the Mississippi are: Division 2, 50%; Division 4, 70%; Division 5, 100%; Division 6, 50%; Division 7, 70%; and Division 8, 80%.

PROBLEM

As a first approximation, the precipitation that falls over an area is partitioned into evaporation, transpiration, changes soil moisture storage, and runoff. This relationship is sometimes called the water balance. Usually we do not have direct measurements of most of the components of the water balance other than precipitation, necessitating the substitution of estimates based on calculations of the water balance.

The Thornthwaite (1948) method for calculating the water balance is commonly applied because it requires only the direct measurement of precipitation and temperature and either the measurement or estimation of the moisture holding capacity of the soil (field capacity). The Thornthwaite method, in common with all models, only approximates the processes in the "real world". It assumes that available moisture is used first for evapotranspiration. The available moisture during any time period is the amount of precipitation during the same time period plus moisture that can be extracted from the soil. Evapotranspiration occurs at a calculated potential rate, a function of the observed mean air temperature, as long as water is not limiting. Water is limiting when the observed precipitation is less than the calculated potential evapotranspiration and the difference cannot be made up by extracting soil moisture. The amount of potential evapotranspiration that cannot be supplied by precipitation plus water extracted from the soil is termed the deficit. When the precipitation is greater than the calculated potential evapotranspiration, the difference first is used to increase the soil moisture. If the the soil moisture is at field capacity or the difference is greater than the amount needed to bring the soil moisture to field capacity, then the excess is termed surplus. From these definitions, it is clear that the Thornthwaite method allows runoff only when there is a surplus, i.e., precipitation is allocated first to evapotranspiration and then to soil moisture storage.

We address three primary questions. The first is the degree to which the computed surplus is related statistically to the observed mean annual runoff of the Mississippi as measured at St. Paul. We know that the relationship will not be one to one because of the assumptions in the Thornthwaite method. The second question is whether the observed fluctuations in the mean annual runoff of the Mississippi at St. Paul are caused by fluctuations in the climate over the drainage basin. Our third question is whether a sharp temperature increase as is often postulated for a doubling of atmospheric carbon dioxide will greatly affect the surplus term in the water balance and, thus, possibly the water resource base of the state.

DATA, METHODS, AND ANALYSES

The data for Minnesota's climatological divisions are available for the average monthly and annual temperature and total monthly and annual precipitation. The available record extends from 1895 to the present, but we use a subset from 1900 to 1982 for reasons set forth below. After 1930 the division data are area averages computed by simply averaging the monthly temperature and precipitation from stations available within the division each month. The number of stations used to determine divisional averages is low in the early part of the record and increases to a maximum in the 1960's. Prior to 1931, the division data are regression estimates based on state-wide averages.

The dependent variable (the variable to be explained) is the mean annual flow of the Mississippi, measured at St. Paul, for calendar years. The data are from the Water Resources Data for Minnesota published by the U.S.G.S. and reported in cubic feet per second. To save storage and help reduce rounding errors in the computations, we divided the flow values by 10 and truncated to four digits. Since this is a linear transformation of interval data, the subsequent computations and analyses are not adversely affected. For presentation, the flow values are reinflated by a factor of ten. We start our analysis with 1899 because few missing monthly river flow measurements occur after 1899, and we needed the first few years of the climate data to bring the Thornthwaite water balance method to equilibrium (suppress the effects of the starting conditions).

The only preliminary analysis of the mean annual flow data was a smoothing by exploratory data techniques (Figure 2). It is clear from the smoothed data that the mean annual flow of the Mississippi at St. Paul declined rather regularly and smoothly from the early 1900's to the late 1930's. Subsequently there was a sharp increase in the mean annual flow in the late 1930's and early 1940's. After the mid-1940's there is little evidence of systematic fluctuation in the mean

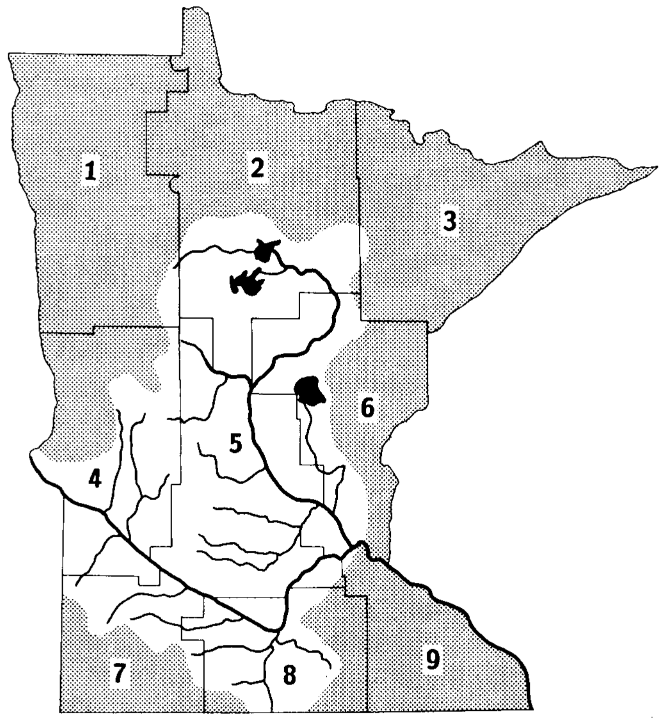

Figure 1. Climatic subdivisions of Minnesota and the Mississippi River Basin above Saint Paul, Minnesota.

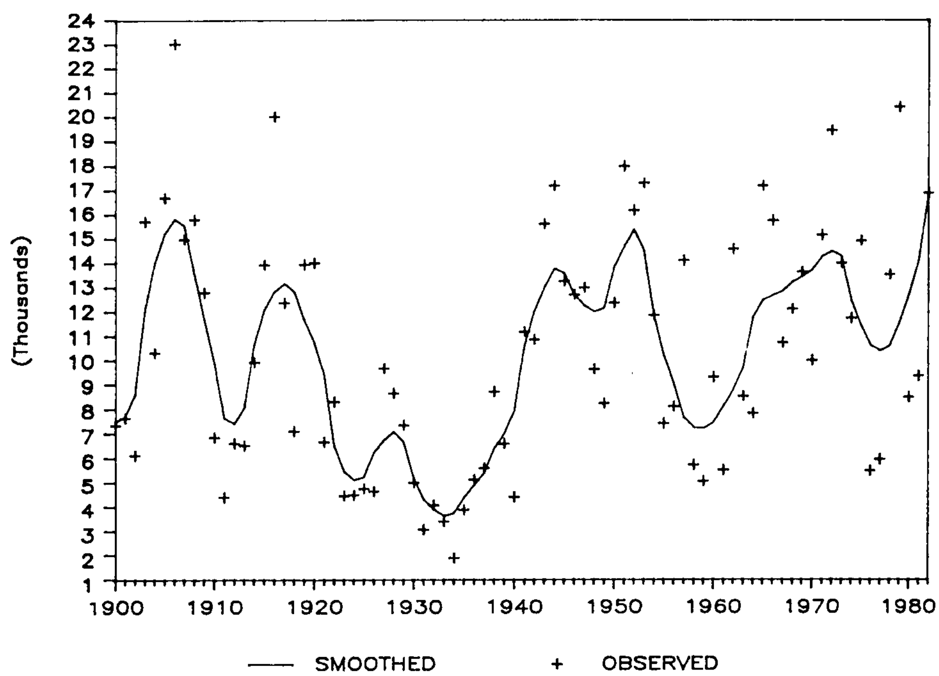

Figure 2. Mean annual flow record of the Mississippi River at Saint Paul, Minnesota.

annual flow time series, although the past few years seem to be strongly upward. Whether the difference between the first and second halves of the record results from climatic fluctuations or other factors is an important question on which we comment later in the analysis.

The monthly division average precipitation and temperature were used to compute the Thornthwaite water budget on a monthly basis. We used the algorithm developed by Willmott, Rowe, and Mintz (1985) because it incorporates a more realistic treatment of snow melt in which the surplus generated by the snow melt is more reasonably placed in the spring rather than spread out over the winter. The computational details will not be given here, but the general procedures are: 1) the potential evapotranspiration is calculated (from observed monthly mean temperatures) for each month and compared with the observed precipitation; 2) soil moisture is either depleted or replenished depending on the difference between precipitation and the calculated potential evapotranspiration and the amount of soil moisture; and 3) actual evapotranspiration, end of month soil moisture quantity, and surplus or deficit (if any) are calculated on a monthly basis. The moisture holding capacity of the soil must be specified as well as an initial value for actual soil moisture. We assume 150 mm for the field capacity in all computations. We started computations in January of 1895 with soil moisture at field capacity and ran the model through December, 1899. This permitted the soil moisture to overcome the arbitrary initial conditions and allowed us to use the computed December soil moisture as the starting value for January, 1900. These calculations were done separately for each of the climatological divisions (2, 4, 5, 6, 7, and 8) of Minnesota contributing to the drainage of the Mississippi at St. Paul.

Our independent variables were created by summing for the calendar year the monthly surpluses, giving an annual total surplus for each division. The time series of annual surpluses for the six climatological divisions are the independent variables in our analyses.

The independent variables are correlated with each other, e.g., it is clear that there is correlation between the surplus time series for divisions 4 and 5 because of spatial proximity. The correlation between independent variables is often called co-linearity. Statistical methods of assessing the relationship between dependent and independent variables, such as multiple linear regression, which is employed here, assume independence of predictor variables. There are two ways to address the problem of co-linearity of independent variables. One method uses stepwise regression in which the predictors are added to the regression equation one at a time in the order specified by the zero order correlation coefficients. Predictor variables are "kept" in the regression equation or "removed" based on a criterion of their contribution to the explained variance of the dependent variable, in this case river flow. The effect is to select a subset of predictor variables that are more or less independent. The disadvantage of the stepwise procedure is that the selected independent variables may be statistically sound but lack a satisfying physical basis. From a physical point of view, it is clear that surpluses from all divisions will contribute to the mean annual flow of the Mississippi at St. Paul, but some divisions will not be included in the statistical model because of co-linearity.

The second method is factor analysis in which the co-linear original variables are transformed into synthetic variables based upon the correlation between the original variables. These new variables are statistically independent and usually a smaller number of variables accounts for the majority of the variance. The disadvantage of factor analysis is that each variable is a linear combination of all of the original variables which makes it very difficult to interpret the components. In particular the first new variable, which accounts for much of the variance, may be an "average" component with substantial contributions from each of the original variables. This problem can be partially overcome by rotation of the principle components to what is called a "simple structure". The simple structure is an attempt to associate one or a very small set of original variables with each principle component. The rotation, in this case, should be orthogonal to retain independence for the statistical analysis. After the principle components are selected the component (or factor) scores can be calculated and used in further analysis. We have used both methods in this study including a varimax rotation of the principle components.

We develop our statistical model based on only the first half of the record. Using only the first half of the record to specify the model has two advantages. First, it allows us to predict the second half of the record and thus determine how well the model works on "independent" data. Second, if the model performs well on the second half of the record, we have evidence that climatic fluctuations are major contributors to the observed differences in the mean annual flow of the Mississippi at St. Paul.

RESULTS

Predicting Mean Annual Flow

The model developed in the stepwise procedure contains a constant and a linear combination of the surpluses in divisions 4 and 6 (Table 1). The coefficient of determination (percentage of variance explained), adjusted for degrees of freedom, is .7 (70%). The mean of the observed annual flow for the full record is 10,418 cfs while the mean for the full record for the estimated (predicted) mean annual flow is 10,383 cfs. The corresponding standard deviations are 4,818 cfs and 4,817 cfs. The model predicts quite well fluctuations in the mean annual flow in both the first and second parts of the record (Figure 3).

A comparison of the first and second halves of the record is revealing. The means of the estimated (predicted) and observed (actual) mean annual flows are identical in the first half of the record (8,699 cfs) while the standard deviations are 4,188 cfs (estimated) and 4,956 cfs (observed). In the second half of the record, the means of the estimated and observed flows are slightly different, 12,027 cfs and 12,097 cfs respectively. The corresponding standard deviations are 4,867 cfs and 4,077 cfs. Thus, the model performs very well on the second half of the record that was not used to develop the model. This is strong evidence that the predictive model is robust and can be used outside of the range of data on the basis of which it was developed.

While the model selected from a stepwise procedure is relatively easy to explain, it is not very satisfactory intellectually because the surpluses from some areas are eliminated to reduce the co-linearity of the independent variables. Thus we also applied a factor and regression analyses to the first half of the record and predicted the full record. Four factors were retained in the analysis and rotated by the varimax criterion (Table 2). The first factor accounts for 33% of the variance and is associated with the surpluses in divisions 7 and 8. The second rotated principle component is associated with surpluses in divisions 6 and 2 and accounts for 32.4% of the variance. Surplus in division 4 is the main contribution to factor 3 which accounts for 23.6% of the variance; factor 4 is a combination of surpluses in divisions 6, 8, and 5, accounting for 9% of the variance. Division 5 surplus contributes about equally to each of the four rotated factors and is not

TABLE 1

STEPWISE REGRESSION MODEL COEFFICIENTS

Variable	Coefficient	T	Probability
Constant	3116.95	3.51	.001
Surplus 4	55.93	4.27	.000
Surplus 6	22.11	2.94	.006
Coefficient of Determination:		.7	
Standard Error of Estimate:		2717.9	

TABLE 2

FACTOR LOADINGS

Variable	Factor(1)	Factor(2)	Factor(3)	Factor(4)
Surplus 7	.833	.346	.407	.075
Surplus 8	.772	.384	.285	.379
Surplus 5	.513	.577	.499	.351
Surplus 2	.338	.872	.327	.112
Surplus 6	.423	.685	.307	.492
Surplus 4	.370	.340	.848	.154

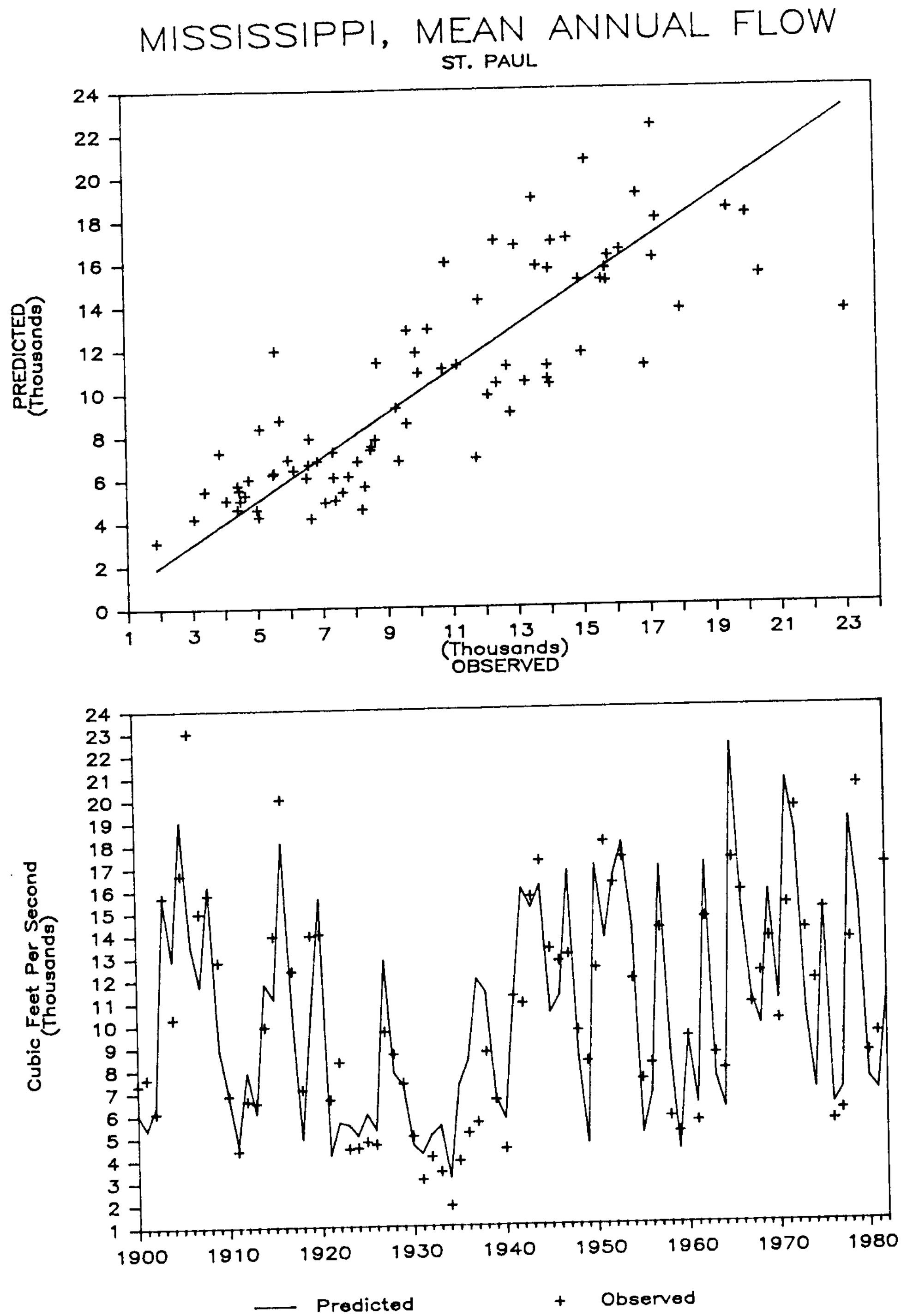

Figure 3. Predicted and observed plots of Mean Annual flow of the Mississippi River at Saint Paul, MN. The upper graph is a plot of the predicted values versus the observed. The lower graph is a plot of both the observed and predicted flows versus time.

dominant in any of the four. It is easy to see from this description that even the rotated factors are combinations of the surpluses which makes it hard to say anything very specific about the affect of any one division. But at least this analysis, unlike the stepwise selected model, makes explicit that surpluses for all of the divisions contribute.

All four rotated factors have significant coefficients in the multiple linear regression model (Table 3). The coefficient of determination is .73 (73% of the variance). The mean of the estimated and observed mean annual flows for the full record are the same, 10,418 cfs, with standard deviations of 4,153 cfs and 4,818 cfs respectively. In the first half the record the mean of the predicted mean flow is about 356 cfs smaller than the observed while it is about 348 cfs larger in the second half of the record. The standard deviation of the predicted mean annual flow is smaller than the observed in both halves.

Additional assessment of these two models can be made by computing error statistics and indices of agreement between the observed and the predicted mean annual flows. Root mean square error statistics are very helpful in assessing the adequacy of the statistical models. Of the two models, the factor analysis model has the lower total root mean square error (Table 4). However, the systematic root mean square error is lower for the model selected by the stepwise procedure (Table 4). In general, lower systematic error is indicative of better model specification, i.e., the errors are likely the result of random error and not the wrong or incomplete selection of independent variables. On this basis, the stepwise model seems better. Other indices of degree of fit of the model are the coefficient of determination between the observed and predicted values and the index of agreement that Willmott (1981) introduced. The factor based model has a slightly larger coefficient of determination, but the index of agreement is much larger for the stepwise based model (Table 4). Thus, we conclude that the stepwise model is better in a predictive sense. However, the factor based model clearly supports the efficacy of the stepwise model.

Flow Response to Climatic Fluctuations

In attempting to answer our second question, whether the increased mean annual flow since the late 1930s is largely the result of climatic fluctuations, we will use only the stepwise regression model results. If we compare the two predictor variables (surpluses in divisions 4 and 6) for the two halves of the record, significant differences occur. For division 4 the mean surplus is 43 mm for the first half and 88 mm for the second half. A t-test of the difference indicates that it is statistically significant. For division 6, the first and second half means are 144 mm and 181 mm respectively. Once again the difference is statistically significant using a t-test. The t-tests were confirmed by non-parametric methods so we are quite confident that we are not making an error in asserting that the amount of surplus available for these two divisions is significantly greater in the second half of the record.

The combination of very good performance by the model in the second half of the record and the significant differences in the predictor variables leads us to conclude that the increase in the mean annual flow in the Mississippi at St. Paul in the last 41 years is largely the result of a fluctuation in the climate as measured by the computed surplus, with much larger surpluses since about 1940. The increase in surpluses as computed by the Thornthwaite method can arise from an increase in precipitation, a decrease in potential evapotranspiration, or a combination of the two. Comparisons between the two halves of the record of the observed

TABLE 3

FACTOR REGRESSION MODEL COEFFICIENTS

Variable	Coefficient	T	Probability
Constant	1041.82	38.08	.0000
Factor 1	2892.99	10.09	.0000
Factor 2	2509.67	8.95	.0000
Factor 3	2754.23	9.69	.0000
Factor 4	1629.62	5.82	.0000
Coefficient of Determination:		.73	
Standard Error of Estimate:		2492.57	

TABLE 4

COMPARISON OF STEPWISE AND FACTOR REGRESSION MODELS

Model Statistics	Stepwise	Factor
Root Mean Square Error	2,759 cfs	2,416 cfs
Systematic Root Mean Square Error	796 cfs	1,217 cfs
Unsystematic Root Mean Square Error	2,642 cfs	2,087 cfs
Coefficient of Determination	0.69	0.74
Index of Agreement	0.91	0.66

precipitations and the computed potential evapotranspirations with t-tests indicates that the computed potential evapotranspirations are not significantly different but that the observed precipitations are. Thus, it would appear that the last 41 years has been a wet period with a significant increase in surplus moisture which has been translated into a significant increase in mean annual flow (and by implication runoff).

Carbon Dioxide Doubling

Most investigations of the consequences of the current rise in carbon dioxide concentration in the atmosphere "predict" a rise of 3 degrees Celsius in the global mean annual temperature. Because temperature can affect surplus by changing the evapotranspiration, we investigated the impact that such a substantial temperature increase might have on Minnesota annual river flow. Our results can only considered a first, and rather crude, estimate because of the many assumptions made. We assume that the increase in Minnesota's temperature will be 3 degrees Celsius, the same as the global average. We further assume that the temperature rise will be concentrated in the winter, and we rather arbitrarily assigned a rise of 4 degrees Celsius in the 6 winter months and a rise of 2 degrees Celsius in the 6 summer months. Finally, we assume that precipitation is not changed.

For the years 1971 through 1983 and divisions 4 and 6, we recomputed the Thornthwaite water budget method using average temperatures increased by these amounts. The results show a drastic decrease in surplus computed (Table 5). In division 4 the average surplus per year changes from 82 mm per year to 36 mm per year, a decrease of 56 percent. The average surplus computed for division 6 decreases by 45 percent from 203 mm to 112 mm per year. Given the importance of the surpluses in these two division to the mean flow of the Mississippi River, it is clear that a major temperature increase would massively affect the water resources of the state.

CONCLUSIONS

The mean annual flow of the Mississippi at St. Paul has been significantly greater in the last 41 years. The increase mean annual flow is closely related statistically to an increase in the water balance surpluses computed for the climatological divisions that contribute to the Mississippi/Minnesota drainage basin in Minnesota. The increases in surpluses are at least partly due to increased divisional average precipitation. The surpluses computed in the Thornthwaite water balance method can be used to estimate (and simulate) water availability at least at this large spatial scale. Finally, it should be noted that surplus in the water budget can be affected by temperature changes. Our crude estimates indicate that temperature rises, such as might occur as a result of increasing carbon dioxide, are large enough to severely decrease the water resource base of the state.

This research establishes the basis for additional work. The excellent performance of the statistical model based on the stepwise regression procedure makes it possible to investigate the statistical properties of the time series of surplus for divisions 4 and 6. If these time series have an autoregressive character, it would be possible to do simulations of mean annual flow of the Mississippi at St. Paul. The results would provide some indication of the limits of expected mean annual flow and the duration of runs of high and low mean annual flow amounts.

TABLE 5

EFFECT OF TEMPERATURE CHANGE ON COMPUTED WATER BALANCE SURPLUS

	Surplus* with Observed Temperature		Surplus* with Increased Temperature		Percent Change**	
Year	Div 4	Div 6	Div 4	Div 6	Div 4	Div 6
1971	178	343	41	167	-67	-51
1972	173	251	136	187	-21	-25
1973	50	198	0	64	---	-69
1974	20	120	0	69	---	-42
1975	103	280	53	172	-49	-39
1976	11	112	0	92	---	-18
1977	31	92	20	52	-35	-43
1978	181	257	89	118	-51	-54
1979	122	242	56	170	-54	-30
1980	42	84	19	48	-55	-43
1981	0	166	0	17	---	-90
1982	74	174	18	146	-76	-18
1983	85	317	37	160	-56	-50
Average	82	203	43	112	-56	-45

*Values in millimeters per year.

**Observed Temp. Surplus/Increased Temp. Surplus*100

A second avenue of future research is attempting to predict the flow in critical months, e.g., July, or the base flow (perhaps using January mean flow as a surrogate) from the surpluses calculated in the Thornthwaite method for the water balance.

It should also be noted that the results of such scenarios simulated here are not confined to river flow, but can reasonably be expected to be felt in the soil moisture and ground-water budgets. Data do not exist to allow us to develop these types of statistical models to test the effects of temperature rise on soil water or ground-water. Physical based simulation models that preserve the continuity of the hydrologic cycle and appropriate geographical information systems file structures to properly handle the data and manage the output maps of such simulations are needed.

REFERENCES

Thornthwaite, C. W., 1948, "An Approach toward a Rational Classification of Climate," Geographical Review, 38, 55.

Willmott, C., 1981, "On the Validation of Models," Physical Geography, 2, 184.

Willmott, C., C. Rowe, and Y. Mintz, 1985, "Climatology of the Terrestrial Seasonal Water Cycle," Journal of Climatology, 5, 589.